

**Economics 8103**  
**Microeconomic Theory**  
**Spring 2005**

**Lecture 8**  
**Subgame Perfect Equilibrium**

I. Introduction

A. Our assumption that it is common knowledge that all players in the game are rational implies that all agents should continue to behave rationally in any state of the game.

1. Moreover, when agents are contemplating possible consequences of various actions, they should anticipate that rational behavior will ensue after any action they might choose.
2. If, in addition, the strategies that are followed in any state of the game are known, as we must assume to apply the Nash equilibrium concept, then the agents should anticipate that after any sequence of actions the subsequent behavior, as described by the strategies, should constitute an equilibrium, at least to the extent that we are able to formulate this principle.
3. The idea of subgame perfect equilibrium is that the behavior in any part of the tree that can be regarded as a game in itself should be a Nash equilibrium of that “subgame.”

B. Historical remarks.

1. Although there are precursors to his work, the formal elaboration of this principle for games of perfect information (every information set contains only a single node) is now attributed to Kuhn (1953).

2. The formalization of this principle for games of imperfect information is due to Reinhard Selten (1965).

C. The outline of the remainder is as follows.

1. We will define the notion of a subgame.
2. The notion of a subgame perfect equilibrium will then be defined.
3. This concept will be illustrated with two examples.
4. We will then state and prove Kuhn's second theorem which describes subgame perfect equilibria of games of perfect equilibrium.
5. Eventually (not in this lecture) we will show that perfect equilibria of the agent normal form (of any extensive form game) are subgame perfect.

## II. The Definition of a Subgame

A. The intuitive description.

1. A subgame must begin at a single node, and it consists of the part of the tree lying below this node.
2. In order for the part of the tree lying below a node to be regarded as a game in itself the agents must know whether they are in the subgame, so the subgame must contain every node in every information set containing any node in the subgame.

B. The formal definition of a subgame.

1. Let an extensive game  $G = ((T, \prec), H, (A, \alpha), (I, \iota), \rho, u)$  of perfect recall be given.
2. For  $t \in T$  let  $T^t = \{t\} \cup \{y | t \prec y\}$ .
  - a. Let  $W^t = \{t\}$ .
  - b. Let  $X^t = X \cap T^t$ .
  - c. Let  $Y^t = \{y | t \prec y\}$ .
  - d. Let  $Z^t = Z \cup T^t$ .

3. **Definition:** We say that  $t$  is the initial node of a subgame if  $H(x) \subset T^t$  for all  $x \in S^t$ .
  - a. Note that this implies that  $\{t\} \in H$  when  $t \in X$ , since perfect recall implies that no information set can contain two nodes one of which precedes the other.
  - b. Note that any  $z \in Z$  is the initial node of a subgame.
4. **Definition:** If  $t$  is the initial node of a subgame then *the subgame beginning at  $t$*  is  $G^t = ((T^t, \prec^t), H^t, (A^t, \alpha^t), (I, \iota^t), \delta_t, u^t)$  where:
  - a.  $\prec^t$  is the restriction of  $\prec$  to  $T^t$ :  $\prec^t = \prec \cap (T^t \times T^t)$ ;
  - b.  $H^t = \{h \in H | h \subset T^t\}$ ;
  - c.  $A^t = \{\alpha(y) | y \in Y^t\}$ ;
  - d.  $\alpha^t: Y^t \rightarrow A^t$  is the restriction of  $\alpha$  to  $Y^t$ ;
  - e.  $\iota^t: H^t \rightarrow I$  is the restriction of  $i$  to  $H^t$ ;
  - f.  $\delta_t \in \Delta(\{t\})$  is the probability measure that assigns all probability to  $t$ .
  - g.  $u^t: Z^t \rightarrow \mathbb{R}^I$  is the restriction of  $u$  to  $Z^t$ .
5. **Remark:** If  $W$  is not a singleton then the game  $G$  is not a subgame of itself, a fact that affects the definitions below in minor ways.

### III. Subgame Perfect Equilibrium

- A. We develop a concept that requires an equilibrium to be an equilibrium in each subgame.
  1. This does not make sense for the normal form, since normal form strategy vectors do not induce behavior in unreached subgames unambiguously.
  2. Therefore we work in the agent normal form, and we begin by defining the restriction of a behavior strategy in the natural way.
  3. We can then define the concept formally.

B. Formal definitions.

1. If  $\pi$  is a behavior strategy for  $G$  and  $t$  is the initial node of a subgame, then  $\pi^t = (\pi_h)_{h \in H^t}$  is the restriction of  $\pi$  to the subgame beginning at  $t$ .
2. **Definition:** A *subgame perfect equilibrium* of  $G$  is a Nash equilibrium  $\pi$  of the agent normal form of  $G$  with the property that  $\pi^t$  is a Nash equilibrium of the agent normal form of  $G^t$  for each subgame  $G^t$ .

IV. Two Examples

A. Consider the game in Figure 1.

[Insert Lecture 8 Figure 1 here.]

1. The agent normal form is

|  |     |        |        |     |
|--|-----|--------|--------|-----|
|  |     | 2      | $U$    | $D$ |
|  | 1   |        |        |     |
|  | $L$ | (1, 2) | (1, 2) |     |
|  | $R$ | (2, 1) | (0, 0) |     |

2. This normal form has two Nash equilibria,  $(L, D)$  and  $(R, U)$ .
3. The only equilibrium of the subgame after  $R$  is for agent 2 to choose  $U$ , so the only subgame perfect equilibrium is  $(R, U)$ .

B. Consider the game in Figure 2.

[Insert Lecture 8 Figure 2 here.]

1. The agent normal form is

|     |           |           |           |     |                          |
|-----|-----------|-----------|-----------|-----|--------------------------|
|     |           | $L$       |           | $R$ |                          |
|     |           | $U$       | $D$       | $U$ | $D$                      |
| $F$ | (3, 0, 3) | (3, 0, 3) | (3, 0, 3) | $F$ | (1, 5, 1)      (5, 7, 5) |
| $B$ | (3, 0, 3) | (3, 0, 3) | (3, 0, 3) | $B$ | (2, 6, 2)      (4, 8, 4) |

2. This game has many Nash equilibria, for instance  $(L, U, F)$ .

3. The agent normal form of the subgame after agent 1 plays  $R$  is the right hand matrix of the agent normal form above. Here  $D$  is a strictly dominant strategy for agent 2, and agent 1's best response to this is  $F$ . The only equilibrium of the entire agent normal form that is compatible with this equilibrium of the agent normal form of the subgame is  $(R, D, F)$ .

## V. Kuhn's Second Theorem

A. A game of perfect information is one in which all decisions are made with complete knowledge of the state of the game.

1. Well known examples are chess, checkers, go, and tic-tac-toe.
2. Those who have any experience with any of these games understand that each position that could occur in one of these games has a value. That is, with "best play" on both sides there is a definite outcome, although there may be several "best moves" in any given position.
3. For this reason it is somewhat misleading to attribute Kuhn's second theorem to Kuhn.
  - a. There are precursors in the mathematical literature, notably a paper by Zermelo.
  - b. I think of Kuhn's contribution as a matter of giving us a precisely formulated understanding of what we knew all along.

B. Formal Definitions.

1. **Definition:** We say that  $G$  is a *game of perfect information* if every information set contains exactly one nonterminal node.
2. **Definition:** We say that  $G$  is *without indifference between outcomes* if there is a space  $\Omega$  of *outcomes* (e.g.  $\Omega = \{\text{win, lose, draw}\}$ ) and a function  $\vartheta: Z \rightarrow \Omega$  such that:
  - a. if  $\vartheta(z) = \vartheta(z')$  then  $u_i(z) = u_i(z')$  for all agents  $i$ ;

- b. if  $\vartheta(z) \neq \vartheta(z')$  then  $u_i(z) \neq u_i(z')$  for all agents  $i$ .
- c. Intuitively we may think of this as saying that if agent  $i$  does not care about the difference between  $z$  and  $z'$ , then no one else does either.

C. The theorem and proof.

**Theorem:** Every game  $G$  of perfect information has a pure strategy subgame perfect equilibrium. If  $G$  is without indifference between outcomes then all subgame perfect equilibria have the same expected payoffs for all agents. (That is, the game has a *value*.) If there is no indifference between terminal nodes (i.e.  $u_i(z) \neq u_i(z')$  for all  $i$  and all distinct  $z, z' \in Z$ ) then there is exactly one subgame perfect equilibrium.

**Proof:** We argue by induction on the *length*  $L(G) = \max_{t \in T} \ell(t)$  of  $G$ . Clearly the result is correct for all games with  $L(G) = 0$ . (These are the degenerate games consisting only of terminal nodes.) Suppose that all claims have been established for all games of perfect information with length less than  $L(G)$ . Then in particular it is true for all subgames of the form  $G^y$  where  $p(y) \in W$ . For each such  $y$  we fix a particular pure strategy subgame perfect equilibrium  $\pi^y$ . This determines an action at all nodes except those in  $W$ , and for each node in  $w$  we let  $\pi_{\{w\}}$  be a (degenerate) probability distribution that assigns all probability to an action in  $A(w)$  that brings about a node whose associated payoff for agent  $\iota(\{w\})$  is at least as large as the payoff associated with any other node in  $F(w)$ . Each selected pure strategy subgame perfect equilibrium  $\pi^y$  leads to a particular terminal node with probability one (since only pure strategies are employed), so although  $\iota(\{w\})$  may be able to choose among several nodes, if the game is without indifference between outcomes then all agents will be indifferent between the nodes that are maximal for him. If the game is without indifference between terminal nodes then, by induction, each subgame  $G^y$  has a unique subgame perfect equilibrium and  $\iota(\{w\})$  will have only one maximizing choice, so there is only one subgame perfect equilibrium for  $G$ .

It remains only to show that the behavior strategy we have constructed is a Nash equilibrium. Of course this is obvious: an information set's expected payoff in the game and in the relevant subgame are monotonically related. (More formal detail on this point is provided in the next lecture.) ■

D. Note that the proof does more than establish existence: we have an algorithmic method for computing a subgame perfect equilibria.

1. That is, one begins by choosing a subgame perfect equilibrium for the subgame beginning at any node with only terminal successors. This allows these nodes to be treated as terminal, so one can iterate this procedure until one arrives at the beginning of the tree.
2. This procedure is called *backward induction* or the *Kuhn algorithm*.
3. This procedure can also be used to prove the existence of subgame perfect equilibria.
  - a. Define a *maximal subgame* to be a proper subgame that is not a proper subgame of another proper subgame.
  - b. Given subgame perfect equilibria of all maximal subgames and a Nash equilibrium, in behavior strategies, of the game obtained by replacing each maximal subgame with a terminal node that has the payoffs resulting from the given equilibrium of the subgame, the behavior strategy obtained from combining all this information is easily shown to be a subgame perfect equilibrium
  - c. This method of proof is not found in books because, for instance, we will prove that sequential equilibria exist, and these are subgame perfect.

E. Subgame perfect equilibrium is generally regarded as completely resolving the philosophical issue of simultaneous rationality in games of perfect infor-

mation.

1. That is, any subgame perfect equilibrium is a mode of simultaneously rational behavior and vice versa.
  2. For this reason this solution concept can be used as a benchmark by which to measure other solution concepts.
- F. If we accept subgame perfect equilibrium as the “right” solution concept for games of perfect information, then we must feel that it is not a complete description of simultaneous rationality in games of incomplete information.
1. To see what I mean by this consider modifying a game of perfect information in the following way:
    - a. Replace the tree with an arborescence consisting of two copies of the original tree.
    - b. The information sets in the new game are the pairs of nodes corresponding to strategic nodes in the original game.
    - c. The initial assessment and the payoffs are chosen so that averaging payoffs across paired terminal nodes according to the initial assessment yields the payoffs of the corresponding terminal node in the original game.
  2. Intuitively we have simply added some ignorance about some “state of the world.”
  3. Since the agents are equally ignorant at the beginning, and no event in the game yields new information about the underlying state, “for practical purposes” the new game is a game of perfect information.
  4. Whereas the original game had many subgames, the new game has none, so subgame perfection does *not* capture our strategic intuitions in the new game.
  5. We will soon see that the sequential equilibrium concept provides a

generalization of subgame perfection that is at least partially successful in addressing these concerns.

II. The example shown in Figure 1 is a simplified version of an example known in the literature as “Rosenthal’s centipede.”

[Insert Lecture 6 Figure 1 here.]

A. This game has the following story.

1. Initially there is \$1 on the table. Agent 1 can either take it, in which case the game is over, or leave it on the table.
2. If agent 1 leaves the dollar then it grows into \$3. Agent 2 can either take all three dollars, in which case the game ends, or he can take two and leave one.
3. If agent 2 leaves a dollar then it grows into three dollars, and agent 1 chooses between taking all three dollars, which ends the game, or taking two and leaving one.
4. This process continues until, in the final period, agent 2 chooses between taking all three dollars or taking two and giving one to agent 1.

B. This is an extensive game of perfect information.

1. Our basic intuition for solving such games is backwards induction.
  - a. We have not discussed this concept formally yet, but it is easy to describe in this example.
  - b. The idea is to start at the bottom of the game tree and work backwards, at each node determining the rational choice in view of the rational choices at nodes lower down in the tree.
2. We execute this “algorithm” concretely.
  - a. If agent 2 is allowed to choose between  $U_3$  and  $D_3$ , rationality requires her to choose  $D_3$ .

- b. Knowing this, if agent 1 chooses between  $L_3$  and  $R_3$ , rationality requires that he choose  $L_3$ .
  - c. Working backwards we find that the other rational choices are  $L_1$ ,  $D_1$ ,  $L_2$ , and  $D_2$ .
  - d. Put informally, common knowledge of rationality requires that each player “kill the goose that lays the golden eggs” before the other player has a chance to. Of course from the players’ point of view this is disastrous.
- C. The logic of backwards induction is quite compelling, but in this example it seems silly and highly unrealistic.
- 1. One proposed resolution is to imagine that there is a small probability that one of the agents is someone who never kills the goose. For any given probability of this, if the number of periods is large one can have “cooperative” equilibria in which the other player keeps the game going by virtue of the combined probability that (a) this type has occurred and (b) the first player will continue to cooperate because he expects the second one to.
  - 2. My own recent thoughts on the centipede are that our instinctive moral feelings about such situations create a divergence between the monetary payoffs and our actual preferences over the possible outcomes.
    - a. Put concretely, wouldn’t it be worth \$1 to not feel like a moronic jerk (this being the consequence of killing the goose on the first opportunity) when the worst that could happen is that you end up feeling that the other person is a moronic jerk.
    - b. It is sometimes suggested that such effects are diminished by making the payoffs large. But suppose the monetary payoffs

were multiplied by 10,000. I think that most people would be even more reluctant to kill the goose right away, fearing a lifetime of regret.

- c. In fact it seems to me quite difficult to construct a version of the centipede in which the actual ordinal preferences over outcomes coincided with the ordering according to monetary outcomes.